Speeding up Switch Markov Chains for Sampling Bipartite Graphs with Given Degree Sequence

Corrie Jacobien Carstens
Korteweg-de Vries Institute for Mathematics, Amsterdam, The Netherlands
c.j.carstens@uva.nl

Pieter Kleer
Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
kleer@cwi.nl

Abstract

We consider the well-studied problem of uniformly sampling (bipartite) graphs with a given degree sequence, or equivalently, the uniform sampling of binary matrices with fixed row and column sums. In particular, we focus on Markov Chain Monte Carlo (MCMC) approaches, which proceed by making small changes that preserve the degree sequence to a given graph. Such Markov chains converge to the uniform distribution, but the challenge is to show that they do so quickly, i.e., that they are *rapidly mixing*.

The standard example of this Markov chain approach for sampling bipartite graphs is the switch algorithm, that proceeds by locally switching two edges while preserving the degree sequence. The Curveball algorithm is a variation on this approach in which essentially multiple switches (trades) are performed simultaneously, with the goal of speeding up switch-based algorithms. Even though the Curveball algorithm is expected to mix faster than switch-based algorithms for many degree sequences, nothing is currently known about its mixing time. On the other hand, the switch algorithm has been proven to be rapidly mixing for several classes of degree sequences.

In this work we present the first results regarding the mixing time of the Curveball algorithm. We give a theoretical comparison between the switch and Curveball algorithms in terms of their underlying Markov chains. As our main result, we show that the Curveball chain is rapidly mixing whenever a switch-based chain is rapidly mixing. We do this using a novel state space graph decomposition of the switch chain into Johnson graphs. This decomposition is of independent interest.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Binary matrix, graph sampling, Curveball, switch, Markov chain decomposition, Johnson graph

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.36

Related Version A full version of this work is available at [5], `https://arxiv.org/abs/1709.07290`.

Funding This work is supported by NWO Gravitation Project NETWORKS, Grant Number 024.002.003.

Acknowledgements We are grateful to Annabell Berger and Catherine Greenhill for some useful discussions and comments on an earlier version of this work. We also thank the anonymous reviewers for their valuable comments.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 36; pp. 36:1–36:18

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The problem of sampling bipartite graphs with a given degree sequence, or equivalently the sampling of binary matrices with fixed row and column sums (marginals) has received a lot of attention in theoretical research, in particular Markov chain Monte Carlo switch-based approaches, see, e.g., [23, 6, 18, 20, 25, 14, 17]. Furthermore, it has many applications, for instance in network science [26] and in the field of ecology [28, 30] where it is used as a null-model. Hence, it is important that the sampling is unbiased and fast. In the Markovian approach, one repeatedly applies small changes to a given binary matrix, that represents the adjacency matrix of a bipartite graph in the natural way, with each change preserving the marginals (the bipartite degree sequence). The idea is that after a sufficient number of changes, i.e., transitions of the underlying Markov chain, the resulting matrix corresponds to an approximately uniform sample from the set of all binary matrices having the given marginals. The number of steps needed to get within a given distance of the uniform distribution is known as the *mixing time* of the Markov chain, and the chain is said to be rapidly mixing if this number can be bounded by a polynomial in the size of the bipartitions considered (the parameters m and n as introduced in Section 4).

The best-known probabilistic procedure for making these small changes uses *switches*, in which two one-entries of the matrix are 'exchanged' with two zero-entries of the matrix see, e.g., [28]. The resulting Markov chain is called the switch chain. A possible implementation for such a chain was proposed by Kannan, Tetali and Vempala [23]. One first chooses two rows and columns uniformly at random, and, if the 2×2 submatrix formed by these rows and columns corresponds to a *checkerboard*, the zero and one entries are exchanged, see Section 4.1 for details. The Curveball algorithm was introduced in [32] and rediscovered in [30]. Both articles focus on applications and are motivated by the need to speed up switch-based algorithms. In the Curveball algorithm one again chooses two rows uniformly at random, but then continues by randomly exchanging all one and zero entries on the chosen rows in a way that preserves the marginals. Such an exchange is called a *(binomial) trade*. A detailed description is given in Section 4.2.

The main result of this work is a theoretical comparison of the relaxation times of the Kannan-Tetali-Vempala (KTV) switch chain, and the Curveball chain. The relaxation time of a Markov chain essentially determines its mixing time. In particular, our result implies that the Curveball chain is rapidly mixing whenever the KTV switch chain is rapidly mixing, but that there is at most a quadratic improvement in relaxation time (Theorem 3). We prove this statement in the more general setting in which there is a set of forbidden entries, that have to be zero. This allows us to also compare the chains for the sampling of a simple directed graph with given degree sequence also, as its adjacency matrix can be given by a square binary matrix with zeros on the diagonal.

In order to establish our results, we introduce a general comparison framework inspired by, and based on, the notion of a heat-bath Markov chain, using the definition by Dyer, Greenhill and Ullrich [12]. This framework essentially compares a given Markov chain with a *locally refined* version, which we will call its *heat-bath variant*. We introduce a novel decomposition of the state space of the switch and Curveball chain in order to apply this framework. In particular, this decomposition allows us to show that the transition matrix of the KTV-switch chain only has non-negative eigenvalues. This result is of independent

interest, as it implies that the chain does not have to be made lazy.[1] Making a Markov chain lazy is an easy way to guarantee that its transition matrix only has non-negative eigenvalues, but the procedure has been described, e.g., as 'unnatural' (see the note of Greenhill [19] for a discussion and references).

As an additional application of our framework, we show it can be used to prove that the parallel Curveball chain [4] is rapidly mixing whenever the Curveball chain is rapidly mixing. In the parallel Curveball chain multiple independent binomial trades are performed in parallel.

Some proofs are omitted from the main text, but can be found in Appendix A.

1.1 Related work

We refer the reader to [17] for a nice exposition on the switch Markov chain. We give a brief overview. Kannan, Tetali and Vempala [23] conjectured that the KTV-switch chain is rapidly mixing for all fixed row and column sums. Miklós, Erdős and Soukup [25] proved the conjecture for half-regular binary matrices, in which all the row sums are equal (or all column sums), and Erdős, Kiss, Miklós and Soukup [14] extended this result to almost half-regular marginals. The authors of [14] prove this in a slightly more general context where there might be certain forbidden edge sets. See also [17, 15] for more results. For rapid mixing results of the switch chain for general (un)directed graphs, see also, e.g., Cooper, Dyer and Greenhill [6], Greenhill [18], and Greenhill and Sfragara [21]. In general, the switch chain can be slow mixing in the presence of forbidden edges sets, which follows from the work of Bezáková, Bhatnagar and Randall [2]. The Curveball algorithm was first described by Verhelst [32] and a slightly different version was later independently formulated by Strona, Nappo, Boccacci, Fattorini and San-Miguel-Ayanz [30]. The name Curveball algorithm was introduced in [30]. The Curveball chain has also been formulated for (un)directed graphs, see Carstens, Berger and Strona [4].

Showing rapid mixing for the switch chain has proven to be a highly non-trivial task [6, 18, 20, 25, 14, 17, 15]. All these works rely directly or indirectly on Sinclair's multi-commodity flow method [29]. In this approach one defines a multi-commodity flow in the state space graph of the switch chain that routes a common amount of flow between any two states of the switch chain. If this can be done in such a way that no edge of the state space graph gets *congested* too severely, then the switch chain is rapidly mixing. The analyses [6, 18, 20, 25, 14] all use the fact that the transition probabilities of the switch chain are polynomially bounded. This property does not hold for the Curveball chain, and therefore one cannot directly use the multi-commodity flows of the switch chain analyses in order to prove that the Curveball chain is rapidly mixing.[2]

Our comparison analysis is a special case of the classical comparison framework developed largely by Diaconis and Saloff-Coste and is based on so-called Dirichlet form comparisons of Markov chains, see, e.g., [7, 8], and also Quastel [27]. See also the expository paper by Dyer, Goldberg, Jerrum and Martin [11]. As the stationary distributions are uniform for all our Markov chains, we can use a more direct, but equivalent, framework based on positive semidefiniteness. We briefly elaborate on this equivalence in Appendix B.

[1] The lazy version of a reversible Markov chain with transition matrix P is the chain with transition matrix $(P + I)/2$, that is guaranteed to only have non-negative eigenvalues. This is mostly done to simplify the use of Sinclair's method [29], one of the most well-known methods for proving rapid mixing of a Markov chain.

[2] This was also briefly mentioned in [4], where the mixing time of the Curveball chain was raised as an open problem.

Finally, the transition matrix of the Curveball Markov chain is a special case of a heat-bath Markov chain under the definition of Dyer, Greenhill and Ullrich [12]. Our work partially builds on [12] in the sense that we compare a Markov chain, with a similar decomposition property as in the definition of a heat-bath chain, to its heat-bath variant.

2 Preliminaries

Let $\mathcal{M} = (\Omega, P)$ be an ergodic, time-reversible Markov chain[3] over state space Ω with transition matrix P and stationary distribution π. We write $P_x^t = P^t(x, \cdot)$ for the distribution over Ω at time step t given that the initial state is $x \in \Omega$. It is well-known that the matrix P only has real eigenvalues $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} > -1$, where $N = |\Omega|$. Moreover, we define $\lambda_* = \max\{\lambda_1, |\lambda_{N-1}|\}$ as the second-largest eigenvalue of P. The *variation distance* at time t with initial state x is

$$\Delta_x(t) = \max_{S \subseteq \Omega} \left| P^t(x, S) - \pi(S) \right| = \frac{1}{2} \sum_{y \in \Omega} \left| P^t(x, y) - \pi(y) \right|,$$

and the mixing time $\tau(\epsilon)$ is then defined as

$$\tau(\epsilon) = \max_{x \in \Omega} \left\{ \min\{t : \Delta_x(t') \leq \epsilon \text{ for all } t' \geq t\} \right\}.$$

A Markov chain is said to be *rapidly mixing* if the mixing time can be upper bounded by a function polynomial in $\ln(|\Omega|/\epsilon)$. It is well-known, e.g., following directly from Proposition 1 [29], that

$$\frac{1}{2} \frac{\lambda_*}{1 - \lambda_*} \ln(1/2\epsilon) \;\leq\; \tau(\epsilon) \;\leq\; \frac{1}{1 - \lambda_*} \cdot (\ln(1/\pi_*) + \ln(1/\epsilon)) \tag{1}$$

where $\pi_* = \min_{x \in \Omega} \pi(x)$. This roughly implies that the mixing time is determined by the *spectral gap* $(1 - \lambda_*)$, or its inverse, the *relaxation time* $(1 - \lambda_*)^{-1}$. Finally, we let $G_\Omega = (\Omega, A)$ be the state space graph, with an arc $(a, b) \in A$ if and only if $P(a, b) > 0$ for $a, b \in \Omega$ with $a \neq b$. If P is symmetric, we define $H_\Omega = (\Omega, E)$ as the undirected counterpart of G_Ω with $\{a, b\} \in E$ if and only if $(a, b), (b, a) \in A$ with $a \neq b$.

2.1 Johnson graphs

One class of graphs that are of particular interest in this work, are the so-called Johnson graphs. For given integers $1 \leq q \leq p$, the undirected Johnson graph $J(p, q)$ contains as nodes all subsets of size q of $\{1, \ldots, p\}$, and two subsets $u, v \subseteq \{1, \ldots, p\}$ are adjacent if and only if $|u \cap v| = q - 1$. We refer the reader to [22, 3] for the following facts. The Johnson graph $J(p, q)$ is a $q(p - q)$-regular graph and the eigenvalues of its adjacency matrix are given by

$$(q - i)(p - q - i) - i \qquad \text{with multiplicity} \qquad \binom{p}{i} - \binom{p}{i - 1}$$

for $i = 0, \ldots, q$, with the convention that $\binom{p}{-1} = 0$. The following observation is included for ease of reference. It will often be used to lower bound the smallest eigenvalue of a Johnson graph.

[3] For more background on Markov chains, we refer the reader to, e.g., [24].

▶ **Proposition 1.** *Let $p, q \in \mathbb{N}$ be given. The continuous function $f : \mathbb{R} \to \mathbb{R}$ defined by*

$$f(x) = [(q-x)(p-q-x)-x] - q(p-q) = x(x-(p+1))$$

is minimized for $x^* = (p+1)/2$, with $f(x^*) = -(p+1)^2/4$.

3 Comparison framework

In this section we describe the comparison framework that will be used to compare the KTV-switch and Curveball Markov chains (Section 4), and to compare the Curveball and the parallel Curveball chain (Section 5). In general, we consider an ergodic (irreducible) Markov chain $\mathcal{M} = (\Omega, P)$ with stationary distribution π, being strictly positive for all $x \in \Omega$, that can be decomposed as[4]

$$P = \sum_{a \in \mathcal{L}} \rho(a) \sum_{R \in \mathcal{R}_a} P_R \tag{2}$$

which is given by a

i) finite index set $\mathcal{L}$, and probability distribution ρ over $\mathcal{L}$,

ii) partition $\mathcal{R}_a = \cup R_{\ell,a}$ of Ω for $a \in \mathcal{L}$,

and where the restriction of a matrix P_R to the rows and columns of $R = R_{\ell,a}$ defines the transition matrix of an ergodic, time-reversible Markov chain on R (and is zero elsewhere), with stationary distribution $\tilde{\pi}_R(x) = \pi(x)/\pi(R)$ for $x \in R$. We use $1 = \lambda_0^R \geq \lambda_1^R \geq \cdots \geq \lambda_{|R|-1}^R$ to denotes its eigenvalues. Note that these are also eigenvalues of P_R and that all other eigenvalues of P_R are zeros (as all rows and columns not corresponding to elements in R only contain zeros). Note that the chain $\mathcal{M}$ proceeds by drawing an index a from the set $\mathcal{L}$, and then performs a transition in the Markov chain on the set R that the current state is in.

The *heat-bath variant* $\mathcal{M}_{heat}$ of the chain $\mathcal{M}$ is given by the transition matrix

$$P_{heat} = \sum_{a \in \mathcal{L}} \rho(a) \sum_{R \in \mathcal{R}_a} \mathbf{1} \cdot \sigma_R \tag{3}$$

with σ_R is a row-vector given by $\sigma_R(x) = \tilde{\pi}_R(x)$ if $x \in R$ and zero otherwise, and $\mathbf{1}$ the all-ones column vector. Intuitively, the chain $\mathcal{M}_{heat}$ proceeds by drawing an index a from $\mathcal{L}$ and then drawing a state x in R with probability $\tilde{\pi}_R(x)$. It can be shown that $\mathcal{M}_{heat}$ is an ergodic Markov chain whenever $\mathcal{M}$ is ergodic, as the state space graph of $\mathcal{M}$ is a subgraph of the state space graph of $\mathcal{M}_{heat}$. It is reversible by construction [12]. The eigenvalues of P_{heat} are always non-negative as was shown in [12].

▶ **Theorem 2.** *Let $\mathcal{M} = (\Omega, P)$ be a Markov chain as in (2) with the property that $\lambda_0^R, \ldots, \lambda_{|R|-1}^R \geq 0$ for all $R \in \mathcal{R}_a$. Let $\mathcal{M}_{heat} = (\Omega, P_{heat})$ be its heat-bath variant as in (3) and let α and β be constants with $\alpha\beta > 0$. If*

$$\alpha - \beta(1 - \lambda_i^R) \geq 0, \tag{4}$$

for all $R \in \mathcal{R}_a$ and $i \in (1, \ldots, |R|-1)$, then P only has non-negative eigenvalues and

$$\frac{1}{\alpha}\frac{1}{1-\lambda_*^{heat}} \leq \frac{1}{\beta}\frac{1}{1-\lambda_*}, \tag{5}$$

where $\lambda_*^{(heat)}$ is the second largest eigenvalue of $P_{(heat)}$. For $\alpha = \beta = 1$, we find $(1-\lambda_*^{heat})^{-1} \leq (1-\lambda_*)^{-1}$.

[4] This description is almost the same as that of a heat-bath chain [12], and is introduced to illustrate the conceptual idea.

The intuition behind Theorem 2 is that in order to compare the relaxation times of a Markov chain and its heat-bath variant, it suffices to compare them locally on the sets R. Note that α and β can both be negative, so that this statement can be used both to upper bound and lower bound the relaxation time of the heat-bath variant in terms of the original relaxation time. The proof of Theorem 2 can be found in Appendix A.

4 Comparing the switch and Curveball chain

In this section we prove our main result as stated in Theorem 3 below. We first introduce some notation and terminology. We are given $n, m \in \mathbb{N}$, fixed row sums $r = (r_1, \ldots, r_m)$, column sums $c = (c_1, \ldots, c_n)$, and a set of forbidden entries $\mathcal{F} \subseteq \{1, \ldots, m\} \times \{1, \ldots, n\}$. We define $\Omega = \Omega(r, c, \mathcal{F})$ as the set of all binary $m \times n$-matrices A satisfying these row and column sums, and for which $A(a, b) = 0$ if $(a, b) \in \mathcal{F}$. The set Ω is the state space of both the switch and Curveball chain. Deciding whether or not Ω is non-empty, and computing an element from it in case it is non-empty, can be done in time polynomial[5] in m and n. That is, in case Ω is non-empty, we can efficiently compute an initial state for the switch or Curveball algorithm. The precise formulations of the transition matrices P_C and P_{KTV} of the Curveball and KTV switch chain, respectively, are given later on in this section.

▶ **Theorem 3.** *Let $r = (r_1, \ldots, r_n)$ and $c = (c_1, \ldots, c_m)$ be given marginals with $n \geq 3$, $\mathcal{F}$ a set of forbidden entries, and assume that $\Omega(r, c, \mathcal{F}) \neq \emptyset$. Let P_C and P_{KTV} be the transition matrices of respectively the Curveball and KTV-switch Markov chains. Then*

$$\frac{2}{n(n-1)} \cdot (1 - \lambda_*^{KTV})^{-1} \;\leq\; (1 - \lambda_*^{C})^{-1} \;\leq\; \min\left\{1, \frac{(2r_{\max}+1)^2}{2n(n-1)}\right\} \cdot (1 - \lambda_*^{KTV})^{-1},$$

where $\lambda_*^{(KTV,C)} = \lambda_1^{(KTV,C)}$ is the second largest eigenvalue of $P_{(KTV,C)}$.

In order to prove Theorem 3, we give a novel decomposition of the state space of the KTV-switch chain. We then show that the Curveball chain is its heat-bath variant. In fact, we introduce a general γ-switch chain, as there are multiple switch-based chains in the literature, and show that the Curveball chain is the heat-bath variant of this general switch chain. The KTV-switch chain corresponds to a specific choice of γ. In the full version [5] we compare the Curveball chain with another switch-based chain for a different value of γ.

4.1 Switch chain

For a given initial binary matrix A, in every step of the Kannan-Tetali-Vempala (KTV) switch algorithm we choose two distinct rows and two distinct columns uniformly at random. If the 2×2 submatrix corresponding to these rows and columns is a *checkerboard* C_i for $i = 1, 2$, where,

$$C_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad C_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then the 2×2 submatrix is replaced by C_{i+1} for i modulo 2 provided the zero entries are not forbidden. That is, if the checkerboard is C_1, it is replaced by C_2, and vice versa. If the

[5] Deciding non-emptiness of Ω can be reduced to deciding if a certain auxiliary graph contains a perfect matching [31]. The latter can be done using the well-known blossom shrinking algorithm [13]. This is also mentioned in [14].

submatrix does not correspond to a checkerboard, nothing is changed. Such an operation is called a *switch*.

Matrices $A, B \in \Omega$ are *switch-adjacent for row i and j* if $A = B$ or if $A - B$ contains exactly four non-zero elements that occur on rows i and j and columns k and ℓ. Two matrices are switch-adjacent if they are switch-adjacent for some rows i and j. In the KTV-switch chain, the probability of transitioning between switch-adjacent matrices is the probability of selecting rows i and j and columns k and ℓ, and always equals $\binom{m}{2}^{-1}\binom{n}{2}^{-1}$.

We need the following additional definitions to introduce the general γ-switch chain. For $A \in \Omega$, we let A_{ij} be the $2 \times n$-submatrix formed by rows i and j, for $1 \leq i < j \leq m$. We define

$$U_{ij}(A) = \{k \in \{1, \dots, n\} : A(i,k) = 1,\ A(j,k) = 0 \text{ and } (j,k) \notin \mathcal{F}\}, \tag{6}$$

with $u_{ij}(A) = |U_{ij}(A)|$, and similarly

$$L_{ij}(A) = \{k \in \{1, \dots, n\} : A(i,k) = 0,\ A(j,k) = 1 \text{ and } (i,k) \notin \mathcal{F}\}, \tag{7}$$

with $l_{ij}(A) = |L_{ij}(A)|$. Note that $L_{ij} \cup U_{ij}$ are precisely the columns k for which A_{ij} has different values on its rows and for which (i,k) and (j,k) are both not forbidden. We will often write u_{ij} and l_{ij} instead of $u_{ij}(A)$ and $l_{ij}(A)$ for brevity.

▶ **Definition 4** (γ-switch chain)**.** Let γ be such that

$$1 - u_{ij}(A)l_{ij}(A) \cdot \gamma > 0 \tag{8}$$

for all $A \in \Omega = \Omega(r, c, \mathcal{F})$ and $1 \leq i < j \leq m$. The transition matrix of the γ-switch chain on state space Ω is given by

$$P_\gamma(A,B) = \begin{cases} \binom{m}{2}^{-1} \cdot \gamma & \text{if } A \neq B \text{ are switch-adjacent,} \\ \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} (1 - u_{ij}(A)l_{ij}(A) \cdot \gamma) & \text{if } A = B, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the transition probability for switch-adjacent matrices is the same everywhere in the state space, and does not depend on the matrices A and B. In particular, the transition matrix P_γ is symmetric and hence the chain is reversible with respect to the uniform distribution. The factor $2/(m(m-1))$ is included for notational convenience. The γ-switch chain can roughly be interpreted as follows. We first choose two distinct rows i and j uniformly at random, and then transition to a different matrix switch-adjacent for rows i and j, of which there are $u_{ij}l_{ij}$ possibilities and where every matrix has probability γ of being chosen; with probability $1 - u_{ij}l_{ij}\gamma$ we do nothing. Taking $\gamma = 2/(n(n-1))$ we obtain the KTV-switch chain [23].

▶ Remark. We always assume the γ-switch chain is irreducible for given r, c and $\mathcal{F}$ (it is clearly always aperiodic, symmetric and finite). Irreducibility is for instance guaranteed in case there are no forbidden entries [28]; or in case $n = m \geq 4$, with $\mathcal{F}$ is the set of diagonal entries and regular marginals $c_i = r_i = d$ for some given $d \geq 1$ [18]. A characterization for irreducibility in the case where $\mathcal{F}$ is the set of diagonal entries is given in [1]. Note that the condition of irreducibility is independent of the value of γ.

4.2 Curveball chain

The Curveball chain proceeds as follows. In every step two rows are chosen uniformly at random from A as in the γ-switch algorithm. Then, a so-called *binomial trade* is performed. In such a trade, we first find all the columns $U_{ij}(A)$ and $L_{ij}(A)$. For example, if $n = 6$ and the $2 \times n$ submatrix is given by

$$\begin{pmatrix} 1 & \mathbf{1} & 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 1 & \mathbf{0} & 0 & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{pmatrix},$$

then we consider the (auxiliary) submatrix

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

given by the second, fourth, fifth and sixth column. We now uniformly at random draw a $2 \times (u_{ij} + l_{ij})$ matrix with columns sums equal to 1, and row sums equal to u_{ij} and l_{ij}. Note that there are $\binom{u_{ij}+l_{ij}}{u_{ij}}$ possible choices, hence the name binomial trade. We then replace the (auxiliary) submatrix with this new submatrix in A. Note that such a drawing can be obtained by uniformly choosing u_{ij} out of $u_{ij} + l_{ij}$ column indices.

Similarly as in the switch chain definition, two matrices A and B are called *trade-adjacent for rows i and j* if $A = B$ or if B can be obtained from A using one binomial trade operation on rows i and j. Two matrices are trade-adjacent if they are trade-adjacent for some row pair.

The Curveball chain $\mathcal{M}_C$ then equals (Ω, P_C) with $\Omega = \Omega(r, c, \mathcal{F})$ and

$$P_C(A, B) = \begin{cases} \binom{m}{2}^{-1} \cdot \binom{u_{ij}+l_{ij}}{u_{ij}}^{-1} & \text{if } A \neq B \text{ are trade-adjacent,} \\ \binom{m}{2}^{-1} \sum_{1 \leq i < j \leq m} \binom{u_{ij}+l_{ij}}{u_{ij}}^{-1} & \text{if } A = B, \\ 0 & \text{otherwise.} \end{cases}$$

4.3 State space decomposition

We next explain how the switch and Curveball chain fit in the comparison framework. The index set $\mathcal{L} = \{(i, j) : 1 \leq i < j \leq m\}$ is the set of all pairs of distinct rows, and ρ is the uniform distribution over $\mathcal{L}$, that is, $\rho(i, j) = \binom{m}{2}^{-1}$ for all $(i, j) \in \mathcal{L}$. The partitions $\mathcal{R}_{(i,j)}$ for $(i, j) \in \mathcal{L}$ are based on the notion of a binomial neighborhood, as defined in [32].

▶ **Definition 5** (Binomial neighborhood)**.** For a fixed binary matrix A and row-pair (i, j), the (i, j)-binomial neighborhood $\mathcal{N}_{ij}(A)$ of A is the set of matrices that can be reached by only applying switches on rows i and j. That is, $\mathcal{N}_{ij}(A)$ contains all matrices that are trade-adjacent to A for rows i and j.

Note that for $B \in \mathcal{N}_{ij}(A)$ we have $U_{ij}(A) \cup L_{ij}(A) = U_{ij}(B) \cup L_{ij}(B)$ and furthermore $u_{ij}(A) = u_{ij}(B)$ and $l_{ij}(A) = l_{ij}(B)$.

Next we will discuss the structure and properties of these binomial neighborhoods. This discussion will culminate into Theorem 7 describing the switch and Curveball chains as being of the forms (2) and (3). Note that we have $A \in \mathcal{N}_{ij}(A)$; if $B \in \mathcal{N}_{ij}(A)$, then $A \in \mathcal{N}_{ij}(B)$ [32]; and, if $A \in \mathcal{N}_{ij}(B)$, $B \in \mathcal{N}_{ij}(C)$, then $A \in \mathcal{N}_{ij}(C)$. That is, the relation $\sim_{ij}$ defined by $a \sim_{ij} b$ if and only if $a \in \mathcal{N}_{ij}(b)$, is an equivalence relation on Ω. The equivalence classes of $\sim_{ij}$ define the sets $\mathcal{R}_{(i,j)}$.

Furthermore, two matrices $A, B \in \Omega$ can be part of *at most* one common binomial neighborhood. This follows directly from the observation that if $B \in \mathcal{N}_{ij}(A) \backslash \{A\}$, then A and

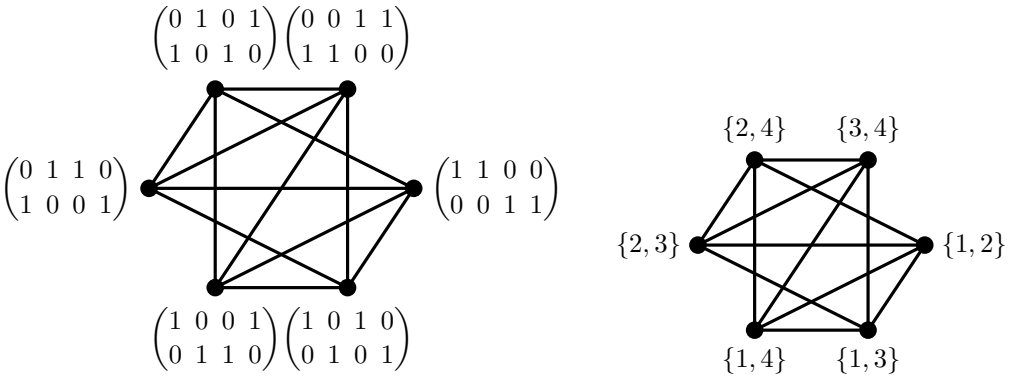

Figure 1 The induced subgraph H for the switch chain on the $(1,2)$-binomial neighborhood of A. On the left we have indexed the nodes by the submatrices of the first four columns, and on the right by label sets, indicating the positions of the 1's on the top row (i.e., row 1).

B differ on precisely rows i and j, so switches using any other pair of rows $\{k,\ell\} \neq \{i,j\}$ can never transform A into B, see [32]. Finally, since $u_{ij}(A) = u_{ij}(B)$ and $l_{ij}(A) = l_{ij}(B)$ when A and B are part of the same binomial neighborhood, these numbers are only neighborhood-dependent, and not element-dependent within a fixed neighborhood. Observe that

$$|\mathcal{N}| = \binom{u_{ij}+l_{ij}}{u_{ij}}.$$

A crucial observation now is that the undirected state space graph H of the γ-switch chain induced on a binomial neighborhood $\mathcal{N}_{ij}$ is isomorphic to a Johnson graph $J(u_{ij}+l_{ij}, u_{ij})$ whenever $u_{ij}, l_{ij} \geq 1$ (see Section 2.1 for notation and definition).[6] To see this, note that every element in the binomial neighborhood $\mathcal{N}_{ij}(A)$ can be represented by the set of indices $U_{ij}(A)$. The set $\{1,\ldots,l_{ij}+u_{ij}\}$ here is then the set of indices of $U_{ij}(A) \cup L_{ij}(A)$. Indeed, matrices $A \neq B$ are switch-adjacent for rows i and j if $U_{ij}(A) \cap U_{ij}(B) = u_{ij} - 1$.

▶ **Example 6.** Consider the binary matrix

$$A = \begin{pmatrix}0&1&1&0&1&0&1\\1&0&0&1&1&0&1\\0&1&0&0&0&1&1\end{pmatrix}$$

and the 2×7-submatrix formed by rows 1 and 2, which is

$$A_{12} = \begin{pmatrix}0&1&1&0&1&0&1\\1&0&0&1&1&0&1\end{pmatrix}.$$

For sake of simplicity, we (uniquely) describe every element of the $(1,2)$-binomial neighborhood $\mathcal{N}_{12}(A)$ by the first four columns (precisely those with column sums equal to one in the submatrix). For the switch chain, the induced subgraph of the undirected state space graph H on the $(1,2)$-binomial neighborhood of A, the Johnson graph $J(4,2)$ is given in Figure 1.

▶ Remark. A fixed binomial neighborhood is reminiscent of the Bernoulli-Laplace Diffusion model, see, e.g., [9, 10] for an analysis of this model. In this model there are two bins

[6] If either $u_{ij} = 0$ or $l_{ij} = 0$ it consists of a single binary matrix.

with respectively k and $n-k$ balls, and in every transition two randomly chosen balls, one from each bin, are interchanged between the bins. Indeed, the state space graph is then a Johnson graph [10]. The transition probabilities are different, due to the non-zero holding probabilities in the switch algorithm, but the eigenvalues of this Markov chain are related to the eigenvalues of the switch Markov chain on a fixed binomial neighborhood, see also [9, 10].

Informally, the Markov chain resulting from always deterministically choosing rows i and j in the switch algorithm, is the disjoint union of smaller Markov chains each with a state space graph isomorphic to some Johnson graph. For a binomial neighborhood $\mathcal{N} = \mathcal{N}_{ij}(A)$ for given $i < j$ and $A \in \Omega$, the undirected graph $H_{\mathcal{N}} = (\Omega, E_{\mathcal{N}})$ is the graph where $E_{\mathcal{N}}$ forms the edge-set of the Johnson graph $J(u_{ij} + l_{ij}, u_{ij})$ on $\mathcal{N} \subseteq \Omega$, and where all binary matrices $B \in \Omega \setminus \mathcal{N}$ are isolated nodes. We use $M(H_{\mathcal{N}})$ to denote its adjacency matrix. The discussion above leads to the following theorem, where we define I_S as the identity matrix on S, and we define J_S as the all-ones matrix on S, that is $J_S(x, y) = 1$ if $x, y \in S$ and zero elsewhere.

▶ **Theorem 7.** *The transition matrix P_γ of the γ-switch chain is of the form (2) namely*

$$P_\gamma = \sum_{1 \le i < j \le m} \binom{m}{2}^{-1} \sum_{\mathcal{N} \in \mathcal{R}_{(i,j)}} \left((1 - u_{ij} l_{ij} \cdot \gamma) \cdot I_{\mathcal{N}} + \gamma \cdot M(H_{\mathcal{N}})\right). \tag{9}$$

The heat-bath variant of the γ-switch chain is given by the Curveball chain, and can be written as

$$P_C = \sum_{1 \le i < j \le m} \binom{m}{2}^{-1} \sum_{\mathcal{N} \in \mathcal{R}_{(i,j)}} \binom{u_{ij} + l_{ij}}{u_{ij}}^{-1} J_{\mathcal{N}}. \tag{10}$$

Proof. The decomposition in (9) follows from the discussion above, and (8) guarantees that the matrix $(1 - u_{ij} l_{ij} \cdot \gamma) \cdot I_{\mathcal{N}} + \gamma \cdot M(H_{\mathcal{N}})$ indeed defines the transition matrix of a Markov chain for every $\mathcal{N}$. Moreover, remember that the γ-switch chain has uniform stationary distribution π over Ω. Indeed, for a binomial neighborhood $\mathcal{N} = \mathcal{N}_{ij}(A)$ for given $i < j$ and $A \in \Omega$, the vector $\sigma_{\mathcal{N}}$ as used in (3) is then given by

$$\sigma_{\mathcal{N}}(x) = \frac{\pi(x)}{\pi(\mathcal{N})} = \frac{1}{|\Omega|} \cdot \frac{|\Omega|}{|\mathcal{N}|} = \frac{1}{|\mathcal{N}|} = \binom{u_{ij} + l_{ij}}{u_{ij}}^{-1}$$

if $x \in \mathcal{N}$, and zero otherwise. This implies that $\mathbf{1} \cdot \sigma_{\mathcal{N}} = \binom{u_{ij}+l_{ij}}{u_{ij}}^{-1} J_{\mathcal{N}}$ as desired. ◀

As a by-product of this decomposition, we now show that the KTV-switch chain [23] only has non-negative eigenvalues when $n \ge 3$. This is of independent interest, as discussed in the introduction.

▶ **Theorem 8.** *The transition matrix of the KTV-switch Markov chain only has non-negative eigenvalues when $n \ge 3$.*

Proof. The KTV-switch chain is exactly the γ-switch chain with $\gamma = 2/(n(n-1))$. As the product $u_{ij}(A) l_{ij}(A)$ can be at most $n^2/4$ for any $A \in \Omega$ and $1 \le i < j \le m$, we see that γ satisfies (8) when $n \ge 3$. To show that P_{KTV} has all non-negative eigenvalues we show that the property assumed in Theorem 2 is satisfied by showing that the matrices

$$Y_{\mathcal{N}} = \left[1 - u_{ij} l_{ij} \cdot \binom{n}{2}^{-1}\right] I_{\mathcal{N}} + \binom{n}{2}^{-1} M(H_{\mathcal{N}})$$

have all non-negative eigenvalues. Theorem 2 then implies that P_{KTV} also only has non-negative eigenvalues. For any eigenvalue λ of this submatrix, we have $\lambda = 1+(\mu-u_{ij}l_{ij})\binom{n}{2}^{-1}$ where $\mu = \mu(\lambda)$ is an eigenvalue of the Johnson graph $J(u_{ij}+l_{ij}, u_{ij})$ on $\mathcal{N}$. In particular, using Proposition 1 with $p = u_{ij}+l_{ij}$ and $q = u_{ij}$, we get $(\mu - u_{ij}l_{ij}) \geq -\frac{1}{4}(u_{ij}+l_{ij}+1)^2 \geq -\frac{1}{4}(n+1)^2$ using that $0 \leq u_{ij}+l_{ij} \leq n$. Therefore, when $n \geq 5$, we have $\lambda \geq 1-(n+1)^2/(2n(n-1)) \geq 0$. The cases $n = 3, 4$ can be checked with some elementary arguments. This is left to the reader. ◀

We conclude this section with the proof of Theorem 3.

Proof of Theorem 3. Let $\mathcal{N} = \mathcal{N}_{ij}(A)$ for given $i < j$ and $A \in \Omega$. Note that the upper bound $(1-\lambda_*^{KTV})^{-1}$ follows from Theorem 2 with $\alpha = \beta = 1$ for which (5) holds as was shown in Theorem 8. We apply Theorem 2 for two pairs (α, β) to obtain the remaining upper and lower bound.

Case 1: $\alpha = 1$ and $\beta = (2n(n-1))/((2r_{\max}+1)^2)$. We show that condition (4) is satisfied. That is, we show that

$$\lambda = 1-\beta\left(1-\left(1+(\mu-u_{ij}\cdot l_{ij})\binom{n}{2}^{-1}\right)\right) = 1+\beta(\mu-u_{ij}\cdot l_{ij})\binom{n}{2}^{-1} \geq 0$$

for any $\mu = \mu(\lambda)$ that is an eigenvalue of the Johnson graph $J(u_{ij}+l_{ij}, u_{ij})$. Again, using Proposition 1 in order to lower bound the quantity $(\mu - u_{ij}\cdot l_{ij})$, we find

$$1+\beta\cdot(\mu-u_{ij}\cdot l_{ij})\binom{n}{2}^{-1} \geq 1-\frac{\beta}{4}(u_{ij}+l_{ij}+1)^2\binom{n}{2}^{-1} \geq 1-\frac{\beta}{4}(2r_{\max}+1)^2\binom{n}{2}^{-1} \geq 0,$$

using the fact that $0 \leq u_{ij}+l_{ij} \leq 2r_{\max}$ and the choice of β. Hence we find the second part of the upper bound.

Case 2: $\alpha = -1$ and $\beta = -(n(n-1))/2$. We have to show that

$$\lambda = \binom{n}{2}\left(1-\left(1+(\mu-u_{ij}\cdot l_{ij})\binom{n}{2}^{-1}\right)\right)-1 = u_{ij}\cdot l_{ij}-\mu-1 \geq 0$$

for all $\mu = \mu(k) = (u-k)(\ell-k)-k$ where $k = 1, \ldots, u$. Note that the eigenvalue $u_{ij}\cdot l_{ij}$ for the case $k = 0$ yields the largest eigenvalue $1 = \lambda_0^{\mathcal{N}}$ of $Y_{\mathcal{N}}$, and does not have to be considered here. The maximum over $k = 1, \ldots, u$ is then attained for $k = 1$, and we have $u_{ij}\cdot l_{ij}-\mu-1 \geq u_{ij}\cdot l_{ij}-((u_{ij}-1)(l_{ij}-1)-1)-1 = u_{ij}+l_{ij}-1 \geq 0$, since $u_{ij}, l_{ij} \geq 1$. This gives us the lower bound and finishes the proof. ◀

5 Parallelism in the Curveball chain

In this section we discuss an additional application of the comparison framework in Section 3. As a binary matrix is only adjusted on two rows at the time in the Curveball algorithm, one might perform multiple binomial trades in parallel on distinct pairs of rows [4]. To be precise, in every step of the so-called *k-Curveball algorithm*, we choose a set of $k \leq \lfloor m/2 \rfloor$ disjoint pairs of rows uniformly at random and perform a binomial trade on every pair (see Section 4.2). For $k = \lfloor m/2 \rfloor$ this corresponds to the Global Curveball algorithm described in [4]. We show that the *k-Curveball chain*, resulting from the k-Curveball algorithm, is the heat-bath variant of the Curveball chain. We use the notation as given in Section 3.

The index set $\mathcal{L} = \mathcal{L}(k)$ is the collection of all sets containing k pairwise disjoint sets of two rows, i.e.,

$$\Big\{\{(1_c, 1_d), (2_c, 2_d), \ldots, (k_c, k_d)\} \ : \ 1_c, 1_d, \ldots, k_c, k_d \in [m],\ |\{1_c, 1_d, 2_c, 2_d, \ldots, k_c, k_d\}| = 2k\Big\},$$

and ρ is the uniform distribution over $\mathcal{L}$. For a fixed collection $\kappa \in \mathcal{L}(k)$, we define the κ-neighborhood $\mathcal{N}_\kappa(A)$ of a binary matrix $A \in \Omega$ as the set of binary matrices $B \in \Omega$ that can be obtained from A by binomial trade-operations only involving the row-pairs in κ. Formally speaking, we have $B \in \mathcal{N}_\kappa(A)$ if and only if there exist binary matrices A_ℓ for $\ell = 0, \ldots, k-1$, so that

$$A_{\ell+1} \in \mathcal{N}_{(\ell+1)_c,(\ell+1)_d}(A_\ell)$$

where $A = A_0$ and $B = A_k$. Note that the matrices A_ℓ might not all be pairwise distinct, as A and B could already coincide on certain pairs of rows in κ. Also note that $u_{i_c i_d}(A) = u_{i_c i_d}(B)$ and $l_{i_c i_d}(A) = l_{i_c i_d}(B)$ if $B \in \mathcal{N}_\kappa(A)$ for $i = 1, \ldots, k$. It is not hard to see that such a neighborhood is isomorphic to a Cartesian product $W_1 \times W_2 \times \cdots \times W_k$ of finite sets[7] $W_1, \ldots, W_k$ with

$$|W_i| = \binom{u_{i_c i_d} + l_{i_c i_d}}{u_{i_c i_d}}.$$

Moreover, the relation $\sim_\kappa$ defined by $a \sim_\kappa b$ if and only if $b \in \mathcal{N}_\kappa(a)$ defines an equivalence relation, and its equivalence classes give the set $\mathcal{R}_\kappa$. We now consider the following artificial formulation of the original Curveball chain: we first select k pairs of distinct rows uniformly at random, and then we choose one of those pairs uniformly at random and apply a binomial trade on that pair. It should be clear that this generates the same Markov chain as when we directly select a pair of distinct rows uniformly at random. For $\mathcal{N}_\kappa \in \mathcal{R}_\kappa$ the matrix $P_{\mathcal{N}_\kappa}$ restricted to the rows and columns in $\mathcal{N}_\kappa$ is then the transition matrix of a Markov chain over $W_1 \times \cdots \times W_k$, where in every step we choose an index $i \in [k]$ uniformly at random and make a transition in W_i based on the (uniform) transition matrix

$$Q_i = \binom{u_{i_c i_d} + l_{i_c i_d}}{u_{i_c i_d}}^{-1} J$$

where J is the all-ones matrix of appropriate size. More formally, the matrix $P_{\mathcal{N}_\kappa}$ restricted to the columns and rows in $\mathcal{N}_\kappa$ is given by

$$\frac{\sum_{i=1}^{k} \left[\otimes_{j=1}^{i-1} \mathcal{I}_j\right] \otimes Q_i \otimes \left[\otimes_{j=i+1}^{k} \mathcal{I}_j\right]}{k}, \tag{11}$$

forming a transition matrix on $\mathcal{N}_\kappa$, and is zero elsewhere. Here $\mathcal{I}_j$ is the identity matrix with the same size as Q_j and $\otimes$ the usual tensor product. The eigenvalues of the matrix in (11) are given by

$$\lambda_{\mathcal{N}_\kappa} = \left\{ \frac{1}{k} \sum_{i=1}^{k} \lambda_{j_i,i} : 0 \le j_i \le |W_i| - 1 \right\} \tag{12}$$

where $1 = \lambda_{0,i} \ge \lambda_{1,i} \ge \cdots \ge \lambda_{|W_i|-1,i}$ are the eigenvalues[8] of Q_i for $i = 1, \ldots, k$. It then follows that

$$P_C = \sum_{\kappa \in \mathcal{L}(k)} \frac{1}{|\mathcal{L}(k)|} \sum_{\mathcal{N}_\kappa \in \mathcal{R}_\kappa} P_{\mathcal{N}_\kappa}$$

---

[7] That is, the elements of W_i describe a matrix on row-pair (i_c, i_d).

[8] See, e.g., [16] for a similar argument regarding the transition matrix, and eigenvalues, of a Markov chain of this form. These statements follow directly from elementary arguments involving tensor products.

which is of the form (2). For $k = 1$, we get back the description of the previous section. Now, its heat-bath variant is precisely the k-Curveball Markov chain

$$P_{k,C} = \sum_{\kappa \in \mathcal{L}(k)} \frac{1}{|\mathcal{L}(k)|} \sum_{\mathcal{N}_\kappa \in \mathcal{R}_\kappa} \frac{1}{|\mathcal{N}_\kappa|} J_{\mathcal{N}_\kappa},$$

where

$$|\mathcal{N}_\kappa| = \prod_{i=1}^{k} \binom{u_{i_c i_d} + l_{i_c i_d}}{u_{i_c i_d}}^{-1}$$

as, roughly speaking, for a fixed neighborhood $\mathcal{N}_\kappa$, the k-Curveball chain is precisely the uniform sampler over such a neighborhood.

▶ **Theorem 9.** *We have* $(1 - \lambda_*^C)^{-1}/k \leq (1 - \lambda_*^{k,C})^{-1} \leq (1 - \lambda_*^C)^{-1}$ *where* $\lambda_*^{k,C}$ *is the second-largest eigenvalue of the* k*-Curveball chain, and* λ_*^C *the second-largest eigenvalue of the original (*1*-)Curveball chain.*

Proof. The upper bound follows from Theorem 2, with $\alpha = \beta = 1$, as the eigenvalues of all the Q_i are non-negative, and therefore (12) implies that the eigenvalues of the matrix in (11) are also non-negative. For the lower bound, we take $\alpha = -1$ and $\beta = -k$. That is, we have to show that $-1 + k(1 - \mu) \geq 0$ with $\mu \in \lambda_{\mathcal{N}_\kappa} \setminus \{1\}$ as in (12). It is not hard to see that the second-largest eigenvalue in $\lambda_{\mathcal{N}_\kappa}$ is $(k - 1)/k$, as the eigenvalues of every fixed Q_i are $1 = \lambda_{0,i} > \lambda_{1,i} = \cdots = \lambda_{|W_i|-1} = 0$. This implies that $-1+k(1-\mu) \geq -1+k(1-(k-1)/k) = 0$ for all $\mu \in \lambda_{\mathcal{N}_\kappa} \setminus \{1\}$. ◀

In general, the upper bound is tight for certain (degenerate) cases, that is, parallelism in the Curveball chain does not necessarily guarantee an improvement in its relaxation time. E.g., take column marginals $c_i = 1$ for $i = 1, \ldots, n$, and row-marginals $r_1 = r_2 = n/2$ and $r_3 = r_4 = 0$, and consider $k = 2$.

6 Discussion

We believe similar ideas as in this work can be used to prove that the Curveball chain is rapidly mixing for the sampling of undirected graphs with given degree sequences [4], whenever one of the switch chains is rapidly mixing for those degree sequences. We leave this for future work, as the proof we have in mind is a bit more involved, but of a very similar nature as the ideas described here.

It should be noted that the main conclusion of our work is not that the Curveball algorithm is necessarily better than the switch-based approaches. In particular, the improvement in relaxation time in Theorem 3, when the maximum row sum is small compared to n, is mostly caused by the fact that the KTV-switch chain is a bad choice of implementation here (as the holding probability of a state in the Markov chain is relatively large in this case). There exist other implementations of the switch chain that are more efficient than the KTV switch chain for certain marginals. For example, the so-called *edge-switch* chain [6, 20]. Here, instead of choosing two rows and columns uniformly at random, one chooses two one-entries of a binary matrix uniformly at random (in order to reduce the probability of staying in the same state of the Markov chain). We give a comparison between these chains in the full version [5]. An interesting direction for future work is to give a better comparison than that in [5]. Although we believe the Curveball chain will outperform any switch-based chain for certain marginals, it not obvious for which marginals this is true. For example, it is not clear to us if this is true

in the case of sampling regular directed graphs with in- and out-degree some small constant. However, for graphs with large regular degrees we expect the Curveball chain to be better.

Moreover, one step of the Curveball algorithm is computationally more expensive than one step of a switch-based algorithm, so although the relaxation time of the Curveball chain might be better than a switch-based chain, this does not automatically imply that the overall running time of the Curveball algorithm is better than that of a switch-based algorithm. Nevertheless, we believe that our results are a first theoretical step for speeding up switch-based Markov chains for sampling bipartite graphs with a given degree sequence.

References

1 Annabell Berger and Matthias Müller-Hannemann. Uniform sampling of digraphs with a fixed degree sequence. In *Graph Theoretic Concepts in Computer Science: 36th International Workshop, WG 2010, Zarós, Crete, Greece, June 28-30, 2010 Revised Papers*, pages 220–231, 2010.

2 Ivona Bezáková, Nayantara Bhatnagar, and Dana Randall. On the Diaconis-Gangolli Markov chain for sampling contingency tables with cell-bounded entries. In *Computing and Combinatorics: 15th Annual International Conference, COCOON 2009 Niagara Falls, NY, USA, July 13-15, 2009 Proceedings*, pages 307–316, 2009.

3 Andries E. Brouwer and Willem H. Haemers. *Spectra of graphs*. Universitext. Springer New York, 2011.

4 Corrie Jacobien Carstens, Annabell Berger, and Giovanni Strona. Curveball: a new generation of sampling algorithms for graphs with fixed degree sequence. *CoRR*, abs/1609.05137, 2016. URL: `http://arxiv.org/abs/1609.05137`.

5 Corrie Jacobien Carstens and Pieter Kleer. Comparing the switch and curveball Markov chains for sampling binary matrices with fixed marginals. *CoRR*, abs/1709.07290, 2017. `arXiv:1709.07290`.

6 Colin Cooper, Martin E. Dyer, and Catherine S. Greenhill. Sampling regular graphs and a peer-to-peer network. *Combinatorics, Probability & Computing*, 16(4):557–593, 2007. `doi:10.1017/S0963548306007978`.

7 Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, 21(4):2131–2156, 10 1993.

8 Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 08 1993. `doi:10.1214/aoap/1177005359`.

9 Persi Diaconis and Mehrdad Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM Journal on Mathematical Analysis*, 18(1):208–218, 1987.

10 Peter Donnelly, Peter Lloyd, and Aidan Sudbury. Approach to stationarity of the Bernoulli-Laplace diffusion model. *Advances in Applied Probability*, 26(3):715–727, 1994.

11 Martin Dyer, Leslie Ann Goldberg, Mark Jerrum, and Russell Martin. Markov chain comparison. *Probab. Surveys*, 3:89–111, 2006.

12 Martin Dyer, Catherine Greenhill, and Mario Ullrich. Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra and its Applications*, 454:57–71, 2014.

13 Jack Edmonds. Paths, trees, and flowers. *Canadian journal of mathematics*, 17(3):449–467, 1965.

14 Péter L. Erdős, Sándor Z. Kiss, István Miklós, and Lajos Soukup. Approximate counting of graphical realizations. *PLOS ONE*, 10(7):1–20, 2015.

15 Péter L. Erdős, Tamás Róbert Mezei, and István Miklós. Efficiently sampling the realizations of irregular, but linearly bounded bipartite and directed degree sequences. *CoRR*, abs/1712.01709, 2017. URL: `https://arxiv.org/abs/1712.01709`.

16 Péter L Erdős, István Miklós, and Zoltán Toroczkai. A decomposition based proof for fast mixing of a Markov chain over balanced realizations of a joint degree matrix. *SIAM Journal on Discrete Mathematics*, 29(1):481–499, 2015.

17 Péter L. Erdős, István Miklós, and Zoltán Toroczkai. New classes of degree sequences with fast mixing swap Markov chain sampling. *CoRR*, abs/1601.08224, 2016. URL: `http://arxiv.org/abs/1601.08224`.

18 Catherine S. Greenhill. A polynomial bound on the mixing time of a Markov chain for sampling regular directed graphs. *Electronic Journal of Combinatorics*, 18(1), 2011.

19 Catherine S. Greenhill. Making Markov chains less lazy. *CoRR*, abs/1203.6668, 2013. URL: `https://arxiv.org/abs/1203.6668`.

20 Catherine S. Greenhill. The switch Markov chain for sampling irregular graphs: extended abstract. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1564–1572, 2015.

21 Catherine S. Greenhill and Matteo Sfragara. The switch Markov chain for sampling irregular graphs and digraphs. *Theoretical Computer Science*, 719:1–20, 2018.

22 Derek A. Holton and John Sheehan. *The Petersen graph*. Cambridge University Press Cambridge, 1993.

23 Ravi Kannan, Prasad Tetali, and Santosh Vempala. Simple Markov-chain algorithms for generating bipartite graphs and tournaments. *Random Structures and Algorithms*, 14(4):293–308, 1999.

24 David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.

25 István Miklós, Péter L. Erdős, and Lajos Soukup. Towards random uniform sampling of bipartite graphs with given degree sequence. *Electronic Journal of Combinatorics*, 20(1), 2013.

26 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002. `doi:10.1126/science.298.5594.824`.

27 Jeremy Quastel. Diffusion of color in the simple exclusion process. *Communications on Pure and Applied Mathematics*, 45(6):623–679, 1992. `doi:10.1002/cpa.3160450602`.

28 A. Ramachandra Rao, Rabindranath Jana, and Suraj Bandyopadhyay. A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā: The Indian Journal of Statistics, Series A*, 58(2):225–242, 1996. `doi:10.2307/25051102`.

29 Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

30 G. Strona, D. Nappo, F. Boccacci, S. Fattorini, and J. San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, 5(4114), 2014.

31 William T. Tutte. The factors of graphs. *Canadian Journal of Mathematics*, 4(3):314–328, 1952.

32 Norman D. Verhelst. An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika*, 73(4):705, 2008.

33 Fuzhen Zhang. *Matrix theory: basic results and techniques*. Universitext. Springer Berlin, 1999.

A Missing proofs

Theorem 2. *Let $\mathcal{M} = (\Omega, P)$ be a Markov chain as in (2) with the property that $\lambda_0^R, \ldots, \lambda_{|R|-1}^R \geq 0$ for all $R \in \mathcal{R}_a$. Let $\mathcal{M}_{heat} = (\Omega, P_{heat})$ be its heat-bath variant as in (3) and let α and β be constants with $\alpha\beta > 0$. If*

$$\alpha - \beta(1 - \lambda_i^R) \geq 0, \tag{4}$$

for all $R \in \mathcal{R}_a$ and $i \in (1, \ldots, |R| - 1)$, then P only has non-negative eigenvalues and

$$\frac{1}{\alpha}\frac{1}{1 - \lambda_*^{heat}} \leq \frac{1}{\beta}\frac{1}{1 - \lambda_*}, \tag{5}$$

where $\lambda_*^{(heat)}$ is the second largest eigenvalue of $P_{(heat)}$. In particular, for $\alpha = \beta = 1$, we find $(1 - \lambda_*^{heat})^{-1} \leq (1 - \lambda_*)^{-1}$.

We will use Propositions 10 and 11 in the proof of Theorem 2. Our proof makes use of positive semi-definite matrices; a symmetric real-valued matrix A is positive semidefinite if all its eigenvalues are non-negative, this is denoted by $A \succeq 0$.

▶ **Proposition 10** ([33]). *Let X, Y be symmetric $\ell \times \ell$ matrices. If $X - Y \succeq 0$, then $\lambda_i(X) \geq \lambda_i(Y)$ for $i = 1, \ldots, \ell$, where $\lambda_i(C)$ is the i-th largest eigenvalue of $C = X, Y$.*

▶ **Proposition 11.** *Let X be the $k \times k$ transition matrix of an ergodic reversible Markov chain with stationary distribution π, and eigenvalues $1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{k-1}$. Let $X^* = \lim_{t\to\infty} X^t$ be the matrix containing the row vector π on every row. Then the eigenvalues of $\alpha(I - X^*) - \beta(I - X)$ are*

$$\{0\} \cup \{\alpha - \beta(1 - \lambda_i) \mid i = 1, \ldots, k - 1\}.$$

for given constants α and β.

Proof. Since X is the transition matrix of a reversible Markov chain, it holds that the matrix VXV^{-1} is symmetric[9], where $V = \text{diag}(\pi_1^{1/2}, \pi_2^{1/2}, \ldots, \pi_k^{1/2}) = \text{diag}(\sqrt{\pi})$. Using the fact that similar[10] matrices have the same set of eigenvalues we determine the eigenvalues of $\alpha(I - X^*) - \beta(I - X)$ by finding those of

$$V(\alpha(I - X^*) - \beta(I - X))V^{-1} = \alpha(I - \sqrt{\pi}^T\sqrt{\pi}) - \beta(I - VXV^{-1}).$$

Let $\mathbf{1} = (1, 1, 1, \ldots, 1)^T$ denote the all-ones vector. We find

$$VXV^{-1}\sqrt{\pi}^T = VX\mathbf{1} = V\mathbf{1} = \sqrt{\pi}^T,$$

so that $\sqrt{\pi}^T$ is an eigenvector of VXV^{-1} with eigenvalue 1. It then follows that $\sqrt{\pi}^T$ is an eigenvector of $\alpha(I - \sqrt{\pi}^T\sqrt{\pi}) - \beta(I - VXV^{-1})$ with eigenvalue 0. Let $\sqrt{\pi}^T = w_0, w_1, \ldots, w_{k-1}$ be a basis of orthogonal eigenvectors for VXV^{-1} corresponding to eigenvalues $\lambda_1, \ldots, \lambda_{k-1}$ (note that X and VXV^{-1} have the same eigenvalues). It then follows that

$$[\alpha(I - \sqrt{\pi}^T\sqrt{\pi}) - \beta(I - VXV^{-1})]w_i = (\alpha - \beta(1 - \lambda_i))\, w_i$$

because of orthogonality. This completes the proof. ◀

[9] This is the same argument that is used to show that a reversible Markov chain only has real eigenvalues.
[10] Two square matrices A and B are *similar* if there exists an invertible matrix T such that $A = T^{-1}BT$.

Proof of Theorem 2. We first show that all eigenvalues of P are non-negative. Let D be the $|\Omega| \times |\Omega|$ diagonal matrix with $(D)_{xx} = \sqrt{\pi(x)}$. Note that the matrices $D^{-1}P_R D$ are positive semi-definite: they are symmetric because P_R defines a reversible Markov chain on R. The eigenvalues of $D^{-1}P_R D$ are equal to those of P_R, which are all non-negative by assumption. Any non-negative linear combination of positive semi-definite matrices is again positive semi-definite, hence $D^{-1}PD \succeq 0$. Thus P has non-negative eigenvalues. A similar argument holds for P_{heat} and was shown in [12]. In particular, this implies that $\lambda_* = \lambda_1$ and $\lambda_*^{heat} = \lambda_1^{heat}$.

Let

$$Y_R := D^{-1}[\alpha(I_R - \mathbf{1} \cdot \sigma_R) - \beta(I_R - P_R)]D$$

where I_R is defined by $I_R(x,x) = 1$ if $x \in R$ and zero otherwise. The matrix Y_R is symmetric since the matrices $\mathbf{1} \cdot \sigma_R$ and P_R define reversible Markov chains on R. Furthermore its eigenvalues are $\{0\} \cup \{\alpha - \beta(1-\lambda_i) \mid i = 1, \dots, k-1\}$ by Proposition 11 and the fact that similar matrices have the same set of eigenvalues. These eigenvalues are non-negative by assumption, hence Y_R is positive semi-definite. It then follows that the matrix

$$D^{-1}[\alpha(I - P_{heat}) - \beta(I - P)]D = \sum_{a \in \mathcal{L}} \rho(a) \sum_{R \in \mathcal{R}_a} D^{-1}[\alpha(I_R - \mathbf{1} \cdot \sigma_R) - \beta(I_R - P_R)]D$$

is also positive semidefinite. Using Proposition 10 and again the fact that similar matrices have the same set of eigenvalues, it follows that

$$\alpha(1 - \lambda_i^{heat}) \geq \beta(1 - \lambda_i)$$

which finishes the proof. ◀

B Markov chain comparison using Dirichlet forms

In this appendix we include some notes on the comparison framework for Markov chains based on Dirichlet forms and show that, for our setting, it is equivalent to a comparison in terms of positive semidefiniteness. The description is taken from Chapter 13.3 [24].

Let $\mathcal{M}$ be an ergodic, reversible Markov chain on state space Ω with transition matrix P and stationary distribution π. The Dirichlet form for the pair (P, π) is defined by

$$\mathcal{E}(f,h) := \langle (I-P)f, h \rangle_\pi$$

for functions $f, h \in \{g \mid g : \Omega \to \mathbb{R}\}$, where $\langle g_1, g_2 \rangle_\pi = \sum_{x \in \Omega} g_1(x) g_2(x) \pi(x)$. To illustrate the usefulness of Dirichlet forms, consider the following result, which appears, e.g., as Lemma 13.22 in [24].

▶ **Lemma 12.** *Let P and $\tilde{P}$ be reversible transition matrices with stationary distributions π and $\tilde{\pi}$, respectively. If $\tilde{\mathcal{E}}(f,f) \leq \alpha \mathcal{E}(f,f)$ for all $f \in \{g \mid g : \Omega \to \mathbb{R}\}$, then*

$$1 - \tilde{\lambda}_1 \leq \left[\max_{x \in \Omega} \frac{\pi(x)}{\tilde{\pi}(x)}\right] \alpha(1 - \lambda_1),$$

where λ_1 and $\tilde{\lambda}_1$ are resp. the second largest eigenvalue of P and $\tilde{P}$. In particular, if both stationary distributions are the same, we get $1 - \tilde{\lambda}_1 \leq \alpha(1 - \lambda_1)$.

The following proposition relates the Dirichlet form to the use of positive semidefinite matrices, in case both stationary distributions are the uniform distribution over Ω. We can then essentially use the above lemma instead of Proposition 10. We choose to give Proposition 10 as this avoids having to introduce the Dirichlet framework.

▶ **Proposition 13.** *Suppose that π and $\tilde{\pi}$ are both the uniform distribution over Ω. Then $\tilde{\mathcal{E}}(f,f) \leq \alpha\mathcal{E}(f,f)$ is equivalent to*

$$\alpha(I-P) \succeq (I-\tilde{P}).$$

Proof. If both stationary distributions are the uniform distribution over Ω, then the condition

$$\tilde{\mathcal{E}}(f,f) \leq \alpha\mathcal{E}(f,f) \tag{13}$$

is equivalent to

$$f^T(I-\tilde{P})f \leq \alpha f^T(I-P)f$$

where the function f is interpreted as a vector. This in turn is equivalent to stating that $\alpha(I-P) \succeq (I-\tilde{P})$. This follows from the equivalence that $A \succeq 0$ if and only if $x^TAx \geq 0$ for all real-valued vectors x. ◀